*D602–D606* *Nucleic Acids Research, 2006, Vol. 34, Database issue*
*doi:10.1093/nar/gkj015*

# CTGA: the database for genetic disorders in Arab populations

**Ghazi O. Tadmouri*, Mahmoud Taleb Al Ali, Sarah Al-Haj Ali and Najib Al Khaja**

Centre for Arab Genomic Studies, PO Box 22252, Dubai, United Arab Emirates

Received July 4, 2005; Revised and Accepted September 17, 2005

## ABSTRACT

**The Arabs comprise a genetically heterogeneous group that resulted from the admixture of different populations throughout history. They share many common characteristics responsible for a considerable proportion of perinatal and neonatal mortalities. To this end, the Centre for Arab Genomic Studies (CAGS) launched a pilot project to construct the 'Catalogue of Transmission Genetics in Arabs' (CTGA) database for genetic disorders in Arabs. Information in CTGA is drawn from published research and mined hospital records. The database offers web-based basic and advanced search approaches. In either case, the final search result is a detailed HTML record that includes text-, URL- and graphic-based fields. At present, CTGA hosts entries for 692 phenotypes and 235 related genes described in Arab individuals. Of these, 213 phenotypic descriptions and 22 related genes were observed in the Arab population of the United Arab Emirates (UAE). These results emphasize the role of CTGA as an essential tool to promote scientific research on genetic disorders in the region. The priority of CTGA is to provide timely information on the occurrence of genetic disorders in Arab individuals. It is anticipated that data from Arab countries other than the UAE will be exhaustively searched and incorporated in CTGA (http://www.cags.org.ae).**

## INTRODUCTION

The Arabs, comprising of 315 million individuals, are living in regions encompassing Mesopotamia, Middle East, Arabian Gulf, North Africa and parts of East and West Africa. In addition, Arab diasporas, with an estimated size of 30 million people, are encountered in all over the world. Although, Arabs consist of heterogeneous groups and many isolates, they share many common characteristics with important influence on their genetic constitution. These include: high rates of inbreeding or consanguineous marriage, elevated birth rates, child bearing in older maternal age and lack of public health measures directed at the control and prevention of congenital and genetically determined disorders (1). All these factors make genetic and congenital disorders responsible for a considerable proportion of perinatal and neonatal mortalities in Arab populations. In fact, congenital malformations are the second leading cause of infant mortality in Bahrain, Kuwait, Oman and Qatar and are the leading cause of infant mortality (40.3%) in the United Arab Emirates [UAE; (2,3)].

Since the 1950s, Arab countries have made progress in medical services leading to better life expectancies and access to health care. Similarly, Arab scholars working in the field of biomedical sciences are giving more attention to publish their results at national or international levels (4). Concurrently, several attempts to review different aspects of genetic diseases in Arab populations were conducted (5–8). However, data were rapidly outdated as new disorders were described in Arabs. To this end, the Centre for Arab Genomic Studies (CAGS) launched a pilot project to construct the 'Catalogue of Transmission Genetics in Arabs' (CTGA) database for genetic disorders in Arabs to educate the medical community and raise public awareness in at-risk populations.

## SOURCE OF INFORMATION

In accordance with its objective to alleviate human suffering from genetic diseases in the Arab World, CAGS coordinated the collection of data on genetic disorders in the UAE population as a model system to be implemented in other Arab countries in the future. Information in CTGA was drawn from two main sources:

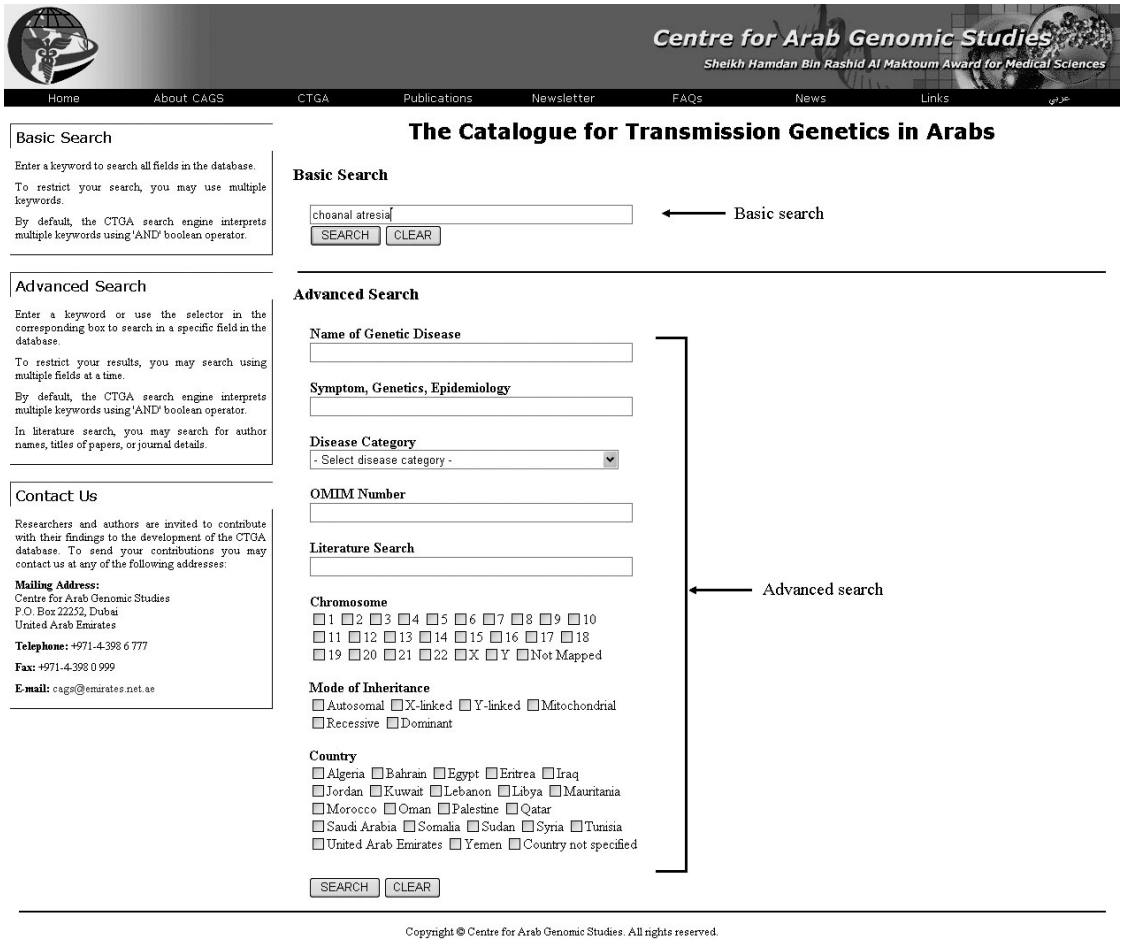
(i) Nationally and internationally published literature: Bibliographic databases were screened for relevant articles on genetic disorders in the UAE. Whenever possible, comprehensive manual scan of hardcopies of national peer-reviewed journals was conducted.
(ii) Laboratory records: In major hospitals of the UAE, patient records covering the last 10–15 years were studied prospectively. These hospitals included laboratories for molecular diagnostics, cytogenetics, biochemistry and

*To whom correspondence should be addressed. Tel: +971 4 398 6 777; Fax: +971 4 398 0 999; Email: tadmouri@hotmail.com, cags@emirates.net.ae

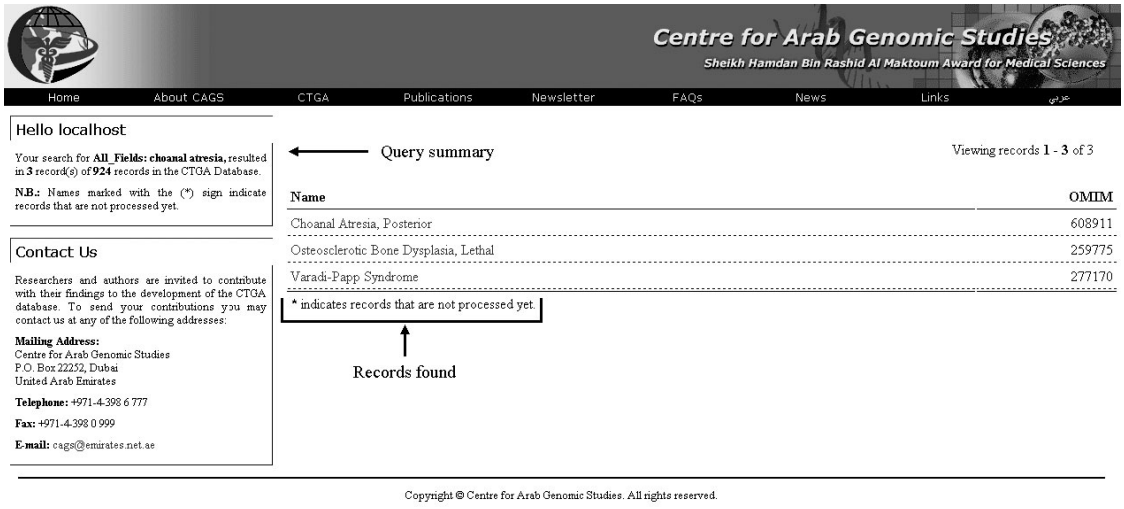
others (1). Detailed information, including the mutation, was recorded using a standardized method. Patient records prove to be an invaluable source of information since they indicated the presence of several inherited disorders for which occurrence data had not been published before.

While data on genetic disorders in patients of various nationalities were collected, only those obtained from UAE nationals and other Arab patients appear in the CTGA database. Furthermore, personal communication with local geneticists provided further insight into the spectrum of inherited disorders in the UAE. Succinctly, the magnitude of genetic disorders and congenital abnormalities reported from the Arab population of the UAE alone (at least 200) demonstrates the efficacy of the algorithm adapted when compared to a review on the subject (9).

(a)




(b)



(c)


**Figure 1.** The online layouts of the CTGA database: (**a**) Search, (**b**) results and (**c**) details.

## THE CTGA DATABASE

The current version of CTGA is a textual database whose structure depends on a web-based search that uses an indexing system for rapid mining of information. As the retrieval of information from the CTGA database is as important as filling data in, we paid considerable attention to provide the users with the option of performing complicated queries to obtain specific results without sacrificing the simplicity. At present, the CTGA database (http://www.cags.org.ae)

can be queried using two modes of search: basic or advanced. In basic search, there is a standard query box in which the user may enter one or more keywords. By default, the CTGA search engine permits the use of wildcards and automatically processes multiple keywords with the 'AND' Boolean operator. However, the power of querying at CTGA lies in its advanced search features. At this point, the user can employ a multitude of user-friendly search combinations according to the name of disease, its classification, symptoms, related gene loci, OMIM number, chromosome location, mode of inheritance, geographic location and others. Proper use of advanced search inevitably increases specificity and narrows down results to a small number of relevant records (Figure 1).

In both types of search, the user issues a command that is interpreted in the CTGA server, processed by the system's language, and results are sent to the user's browser as a standard HTML document with no requirement for any additional software. Query results are alphabetically listed in table form and include the names and corresponding OMIM numbers of genes and genetic disorders described in the Arab people. By selecting a name in the table of results, the user is able to access extensive details relating to a specific gene or genetic disorder (Figure 1).

A detailed record includes text-, URL- and graphic-based fields. The title and alternative names indicate the primary title and alternative titles and symbols of the disorder or gene. A graphical map demonstrates the geographical origin of the individuals described in the entry. A disorder is categorized according to the World Health Organization International Classification of Disease (WHO-ICD) $10^{th}$ revision. OMIM number is a URL-based field that takes the user to the corresponding file of the gene or disorder at the Online Mendelian Inheritance in Man (OMIM) database (10). Information regarding Gene Map Locus is drawn primarily from OMIM. Mode of Inheritance, Description and Molecular Genetics are textual fields that contain summaries on the clinical features and genetic pathology for the corresponding entry. Epidemiology in the Arab World is the major part of an entry since it includes a detailed review of research analyses regarding the gene loci or clinical phenotypes in Arab individuals. References within an entry are linked to their corresponding PubMed abstracts except for articles from national peer-reviewed medical journals not indexed in PubMed. Following the references are two URL-based fields. Related CTGA Records takes the user to any intra-CTGA entry(ies) with a shared relationship(s) while Links anchors at external resources with additional information. Authors who contribute with additions or changes to the entry are given credit in the Contributors field along with the date when the contribution was submitted. Changes made by the editorial staff are documented in the Edit History field (Figure 1).

As of October 1 2005, CTGA had 692 phenotype entries and 235 related gene entries with descriptions in Arab individuals. In the UAE, CTGA information includes about 213 phenotypic descriptions (including 14 in Arab non-UAE nationals) and 22 related genes (including 3 in Arab non-UAE nationals). Currently, authors at the Centre for Arab Genomic Studies create about 25 entries and update an equivalent number each month. Although CTGA has a short lifespan on the public domain of the Internet, it averages at least 150 unique users per day. The peak of simultaneous users accessing the database usually occurs between 03:00 and 12:00 GMT.

## SIGNIFICANCE OF CTGA

### A tool for decision-making in health-related domains

The geographical distributions of genetic disorders in CTGA can either be restricted to small locales (Stuve–Wiedemann syndrome), commonly widespread (beta-thalassemia), or reflect a patchy distribution (alpha-thalassemia) although a high prevalence is expected in the region (1). On the other hand, the molecular/biochemical pathologies in ~25% of genetic disorders described in Arabs have not been determined yet, thus, these serve as excellent candidates for linkage analyses and genotype/phenotype studies (1). Obviously, the interpretation of these data is an important tool for authorities to decide on future health-related strategies and to propose research directions on disorders for which information is still scant.

### A hub of locally produced scientific information on genetic disorders

Current data on genetic disorders in CTGA reflect a bias in the geographical distribution. At present, genetic disorders recorded in Tunisia, Lebanon, Morocco and Saudi Arabia add up to ~40% of all genetic disorders in Arab populations. The main reason for this is the well-established custom of scientific reporting at international level, while in other Arab countries reports mostly appear in national publications or stay confined to non-public laboratory records (11). Consequently, by publishing scientific information locally produced in peer-reviewed journals and unpublished data collected from records of laboratories from the Arab World, CTGA exposes valuable local information that is not accessible to the large scientific community.

### A catalyst for establishing collaborations with Arab scientific groups

The extended consanguineous family structure, commonly present in Arab societies, is an important factor leading to the propensity of severe congenital inherited diseases in most Arab populations (1). Incidentally, genetic disorders in Arabs tend to display peculiar distribution patterns not present in many other world populations. A major model that explains this concept is the vertical dissemination of a genetic mutation in an Arab family, where mutation carriers mostly remain concentrated within the extended family; thus, offering great opportunities to depict the genetic nature of their disease predisposition (1). In view of all the above, the wealth of information that CTGA is accumulating is, in our opinion, an indispensable tool for scientists to recognize Arab colleagues working on similar domains and decide on possible collaborations or exchange of know-how.

### An educational tool on genetic disorders in the region

Studies have clearly indicated that the correct dissemination of knowledge is an important step towards the eradication of genetic disorders in Arab populations (12). The CTGA

database plays such an educational role as it addresses both the medical communities and at-risk populations.

## FUTURE OF CTGA

The priority of CTGA is to provide timely information on the occurrence of genetic disorders in the Arab populations. It is anticipated that data from Arab countries other than the UAE will soon be exhaustively searched and incorporated in CTGA. A similar strategy to that applied in the UAE will be adapted using published and unpublished information. In this regard, CAGS is currently forming a council of Arab geneticists to orchestrate data collection from their corresponding countries. Besides, all geneticists working on genetic disorders in Arabs are hailed to contribute to the growth of CTGA.

On the other hand, CAGS is planning to develop CTGA into a data mart that collects information from different specialized databases using relational data models of database management systems (DBMS). These specialized databases may include DNA sequence data, mutations, polymorphisms or disorder-specific information. Certainly, such integration will add new benefits and uses of CTGA, the credence of which is simply defined by the feedback received from the database users.

## ACKNOWLEDGEMENTS

The Centre for Arab Genomic Studies is a division of H.H. Sheikh Hamdan Bin Rashid Al Maktoum Award for Medical Sciences. Funding is provided by Sheikh Hamdan Bin Rashid Al Maktoum Award for Medical Sciences. We thank the Executive Committee Members of CAGS and colleagues in the various medical and academic centers who facilitated data collection in the UAE. The authors wish to thank Dr Erol Baysal for reviewing the material for this manuscript and for providing insightful comments and suggestions. Funding to pay the Open Access publication charges for this article was provided by the Centre for Arab Genomic Studies.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Tadmouri,G.O., Al Ali,M.T. and Al Khaja,N. (2004) *Genetic Disorders in the Arab World: United Arab Emirates*. Centre for Arab Genomic Studies, Dubai, UAE.
2. Hamamy,H. and Alwan,A. (1994) Hereditary disorders in the Eastern Mediterranean Region. *Bull. World Health Organ.*, **72**, 145–154.
3. Al Hosani,H. and Czeizel,A.E. (1996) Unique demographic situation in the United Arab Emirates. *Am. J. Med. Genet.*, **61**, 1.
4. Tadmouri,G.O. and Bissar-Tadmouri,N. (2003) Biomedical publications in an unstable region: the Arab world, 1988–2002. *Lancet*, **362**, 1766.
5. El-Hazmi,M.A.F. and Warsy,A.S. (1996) Genetic disorders among Arab populations. *Saudi Med. J.*, **17**, 108–123.
6. Alwan,A., Modell,B., Czeizel,A. and Hamamy,H. (1997) *Community Control of Genetic and Congenital Disorders*. WHO EMRO Technical Publications Series No. 24, Alexandria, Egypt.
7. Teebi,A.S. and Farag,T.I. (1997) *Genetic Disorders Among Arab Populations*. Oxford University Press, UK.
8. Abel,E.L. (2003) *Arab Genetic Disorders: A Layman's Guide*. McFarland & Company, NC.
9. Al Talabani,J., Shubbar,A.I. and Mustafa,K.E. (1998) Major congenital malformations in United Arab Emirates (UAE): need for genetic counselling. *Ann. Hum. Genet.*, **62**, 411–418.
10. Hamosh,A., Scott,A.F., Amberger,J.S., Bocchini,C.A. and McKusick,V.A. (2005) Online mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
11. Tadmouri,G.O. (2004) Biomedical science journals in the Arab world. *Saudi Med. J.*, **25**, 1331–1336.
12. Khlat,M., Halabi,S., Khudr,A. and Der Kaloustian,V.M. (1986) Perception of consanguineous marriages and their genetic effects among a sample of couples from Beirut. *Am. J. Med. Genet.*, **25**, 299–306.